

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ  
СІКОРСЬКОГО»**

**ДУБІНІНА СВІТЛАНА ВІТАЛІЇВНА**



УДК 004.852

**БАЙЄСІВСЬКІ МЕТОДИ МОДЕЛЮВАННЯ АКТУАРНИХ ПРОЦЕСІВ ТА  
ОЦІНЮВАННЯ РИЗИКІВ СТРАХОВИХ КОМПАНІЙ**

**05.13.23 – Системи та засоби штучного інтелекту**

**АВТОРЕФЕРАТ  
дисертації на здобуття наукового ступеня  
кандидата технічних наук**

**Київ – 2017**

**Дисертацією є рукопис.**

Роботу виконано у Навчально-науковому комплексі “Інститут прикладного системного аналізу” Національного технічного університету України “Київський політехнічний інститут імені Ігоря Сікорського”.

**Науковий керівник:** доктор технічних наук, професор  
**Бідюк Петро Іванович**,  
Навчально-науковий комплекс “Інститут прикладного системного аналізу” Національного технічного університету України “Київський політехнічний інститут імені Ігоря Сікорського”, професор кафедри математичних методів системного аналізу.

**Офіційні опоненти:** доктор технічних наук, професор  
**Корабльов Микола Михайлович**,  
Харківський національний університет радіоелектроніки, професор кафедри електронних обчислювальних машин.

кандидат технічних наук, старший науковий співробітник

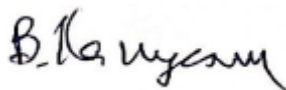
**Пономаренко Сергій Олексійович**, Державний науково-дослідний інститут авіації, провідний науковий співробітник.

Захист відбудеться 4 квітня 2017 р. о 15 годині 00 хвилин на засіданні спеціалізованої вченої ради Д 26.002.03 при Національному технічному університеті України “Київський політехнічний інститут імені Ігоря Сікорського” за адресою: 03056, Київ, просп. Перемоги, 37, корп. 35, ауд. 001.

З дисертацією можна ознайомитися у бібліотеці Національного технічного університету України “Київський політехнічний інститут імені Ігоря Сікорського” за адресою: 03056 м. Київ, пр. Перемоги, 37.

Автореферат розіслано 2 березня 2017 р.

Вчений секретар  
спеціалізованої Вченої ради Д 26.002.03



В. О. Капустян

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** У зв'язку з удосконаленням технологій пошуку та обробки даних у різних сферах фінансово-економічної діяльності накопичуються великі масиви даних різної природи і типів. Так, функціонування підприємств усіх видів власності супроводжується реєстрацією і записом відповідної докладної інформації стосовно їх діяльності, а також про споживачів товарів і послуг. У результаті, величезні масиви даних досить часто утворюють хаотичні нагромадження, які без спеціальної обробки є непотрібними; більше того, вони ускладнюють процес пошуку необхідної інформації. Критерії ефективної обробки таких даних повинні враховувати їх великі обсяги та різноманітність, а результати повинні бути конкретними та зрозумілими; інструменти для обробки повинні бути простими, адаптованими до користувачів різних рівнів та зручними у використанні.

Для високоякісної обробки статистичних і експериментальних даних необхідно застосовувати як класичні статистичні методи, так і сучасні методи інтелектуального аналізу даних (ІАД). Методи ІАД виникли та набули подальшого розвитку на базі досягнень прикладної статистики, функціонального моделювання, розпізнавання образів, методів штучного інтелекту, теорії баз даних тощо. Встановлено, що існують вироджені вибірки статистичних даних, які потребують докладного аналізу з метою розв'язання задач коректної обробки екстремальних значень, зокрема у сфері актуарного моделювання. Після попередньої обробки такі дані використовують для побудови математичних моделей, оцінювання прогнозів та можливих ризиків фінансових втрат. ІАД полягає у виявленні у первинних даних раніше невідомої, доступної, практично корисної і нетривіальної інтерпретації знань, необхідної для прийняття рішень у різних сферах діяльності. До методів і алгоритмів ІАД належать такі: нейронні мережі, дерева рішень, нечітка логіка та нейро-нечіткі системи, символічні правила, методи ближчого сусіда, метод опорних векторів, байєсівські мережі та байєсівська регресія, ієрархічні та неієрархічні методи кластерного аналізу, еволюційне програмування і генетичні алгоритми, різноманітні методи візуалізації даних та ін. Висока популярність методів ІАД зумовлена їх гнучкістю, багатофункціональністю, можливістю боротьби з невизначеностями різних типів і природи, можливістю побудови багатовимірних комбінованих моделей з високими характеристиками якості оцінок прогнозів та керування на їх основі.

Широкого застосування ІАД набув у розв'язуванні актуарних задач, оскільки інструменти візуалізації надають можливість виконувати аналіз даних предметними фахівцями, які не володіють відповідними математичними знаннями. ІАД являють собою велику цінність для керівників і актуаріїв у сфері страхування. За допомогою методів ІАД вони можуть одержати відчутні переваги у конкурентній боротьбі шляхом розв'язання таких задач: прогнозування страхових виплат у разі настання страхового випадку; оцінювання фінансових ризиків страхових компаній для недопущення подальшого банкрутства; виявленню шахрайства та ін. Таким чином виникають задачі аналізу та менеджменту фінансових ризиків з використанням сучасних методів математичного моделювання, прогнозування та ефективної

підтримки прийняття рішень із застосуванням методів інтелектуального аналізу даних.

Актуальність даного дослідження полягає у застосуванні сучасних методів ІАД до моделювання і прогнозування розвитку актуарних процесів та оцінювання фінансових ризиків страхових компаній (СК). Це пояснюється тим, що сьогодні сучасний стан фінансів СК потребує пошуку нових форм та методів підвищення їх конкурентоспроможності, удосконалення та впровадження у практику економіко-математичних моделей для оцінювання і прогнозування їх подальшого розвитку в умовах значної конкуренції та невизначеності, ризику реалізації багатьох процесів, які зустрічаються фізичним особам та підприємствам різних форм власності у їх діяльності. Розмаїття форм прояву ризику, частота та складність наслідків його реалізації зумовлюють необхідність поглибленого аналізу можливих ризиків та економіко-математичного обґрунтування фінансової політики СК. Використання саме економіко-математичних методів для аналізу та управління фінансовими ризиками обумовлене їх можливістю отримання обґрунтованої та достовірної оцінки основоположних характеристик фінансової стабільності. Отже, проблема створення систем для ефективнішого оцінювання фінансового стану з метою уникнення і недопущення банкрутства СК та підвищення рівня її фінансової стабільності є надзвичайно актуальною.

Сучасні розробки стосовно створення системного підходу, спрямованого на виконання поглибленого аналізу складних систем і, зокрема, економіко-математичних явищ, належать академіку Згуровському М. З., професорам Панкратовій Н. Д., Касьянову П. О., Капустяну В. О., Романенку В. Д., Зайченку Ю. П., Снитюку В. Є., Макаренку О. С., Данилову В. Я., Поладчикову В. М. та зарубіжним вченим: Е. Сейджу, Дж. Мелса, Н. Хастінгсу, Дж. Пікоку та іншим. Значний вклад у теорію і практику актуарних досліджень внесли такі відомі вчені: Х. У. Гербер, Дж. Ч. Хікман, С. Дж. Несбітт, Н. Л. Бауерс, С. С. Іванов, В. В. Шахов, М. Денуйт та інші.

Для аналізу та оцінювання прогнозів динаміки розвитку актуарних процесів у дисертаційній роботі запропоновано методику моделювання і створення нових математичних моделей на основі використання структур узагальнених лінійних моделей (УЛМ). Всі етапи реалізації цієї методики супроводжуються використанням відповідних множин статистичних критеріїв, що сприяє підвищенню якості проміжних та остаточних результатів обчислень. Для ілюстрації застосування запропонованої методики зібрано необхідні фактичні статистичні дані. В результаті попереднього аналізу наявної статистичної вибірки виявлено підмножину вироджених даних з екстремальними значеннями. Для подальшого опрацювання цих даних розв'язано задачу інтелектуальної обробки екстремальних значень з метою наближення вибірки до розподілу ймовірностей визначеного типу. Оцінювання невідомих параметрів математичних моделей виконано із застосуванням методів максимальної правдоподібності та байєсівського підходу, які відносяться до інтелектуального аналізу даних.

### **Зв'язок роботи з науковими програмами, планами, темами**

Дисертаційна робота виконана у навчально-науковому комплексі «Інститут прикладного системного аналізу» відповідно до планів науково-дослідних робіт кафедри математичних методів системного аналізу: д/б НДР «Розробка інформаційної технології моделювання та оцінювання фінансово-економічних ризиків із врахуванням невизначеностей різної природи (на основі байєсівських моделей)» (№ ДР: 0113U000650); д/б НДР «Розробка методології системного аналізу, моделювання та оцінювання фінансових ризиків» (№ ДР: 0115U000356).

### **Мета і завдання дослідження**

Метою дослідження є розробка методології побудови адаптивних математичних моделей для прогнозування та оцінювання ризиків актуарних процесів за допомогою узагальнених лінійних моделей та байєсівського підходу; підвищення адекватності моделей актуарних процесів і якості оцінок прогнозів виплат внаслідок настання страхових випадків та операційних втрат страхових компаній.

Для досягнення мети дослідження необхідно поставити та розв'язати такі наукові завдання:

1. Виконати аналіз сучасних методів моделювання та прогнозування актуарних ризиків у сфері страхування та факторів, що впливають на страхові виплати.
2. Удосконалити методику побудови математичної моделі для опису та прогнозування актуарних процесів із екстремальними даними.
3. Побудувати моделі, що належать до двох класів математичних моделей для аналізу та прогнозування величини виплат у сфері страхування.
4. Розробити та апробувати модель у формі мережі Байєса для оцінювання величини операційного ризику СК та запропонувати заходи щодо його уникнення.
5. Проаналізувати якість побудованих моделей та формування висновків на їх основі щодо можливості їх застосування для розв'язання задачі прогнозування та оцінювання ризику актуарних процесів.
6. На основі системного підходу побудувати та реалізувати систему підтримки прийняття рішень (СППР) для розв'язання задач моделювання та прогнозування величини страхових виплат при настанні страхового випадку і оцінювання операційного ризику страхових компаній.

*Об'єкт дослідження* – актуарні процеси і технології інтелектуального аналізу даних у страхуванні в умовах наявності невизначеності та ризику.

*Предмет дослідження*: байєсівські методи моделювання та оцінювання невідомих параметрів моделей, а саме: узагальнених лінійних моделей і байєсівських мереж для розв'язання задачі моделювання актуарних процесів та оцінювання операційних ризиків.

**Методи дослідження**: теорія узагальнених лінійних моделей та екстремальних значень, байєсівські мережі і теорія формування імовірнісного висновку, методи оцінювання невідомих параметрів математичних моделей (Монте-Карло і метод максимальної правдоподібності).

**Наукова новизна отриманих результатів** визначається такими теоретичними і практичними результатами, отриманими автором:

*Вперше:*

- запропоновано метод моделювання актуарних процесів та оцінювання ризиків в умовах виродженості початкової вибірки, який відрізняється від відомих урахуванням екстремальних значень і забезпечує підвищення якості оцінок прогнозів за узагальненими лінійними моделями;
- побудовано та досліджено узагальнені лінійні моделі для формального опису актуарних процесів, які відрізняються модифікованою комплексною структурою і забезпечують можливість високоякісного оцінювання операційного ризику;
- побудовано нову модель у формі байєсівської мережі для оцінювання операційних ризиків у страхуванні, яка відрізняється коректністю формального опису за байєсівським інформаційним критерієм і забезпечує обчислення високоякісних оцінок прогнозів розвитку актуарних процесів.

*Удосконалено:*

- метод оцінювання екстремальних значень, який відрізняється комплексним застосуванням теорії екстремальних значень та байєсівського підходу і забезпечує подолання проблеми виродженості вибірки статистичних даних.

*Отримали подальший розвиток:*

- розроблено системну методологію побудови адаптивних моделей актуарних процесів, яка відрізняється запропонованим методом моделювання в умовах виродженості вибірки, методом оцінювання екстремальних значень і забезпечує підвищення адекватності моделей та якості оцінок прогнозів.

### **Практичне значення одержаних результатів**

1. Усі теоретичні розробки, отримані автором, доведені до конкретних моделей, інженерних методик, алгоритмів та засобів прогнозування виплат страхової компанії. Вони забезпечують функціонування СППР в умовах її постійної адаптації до ризикових ситуацій в умовах невизначеності (а саме: наявності екстремальних значень, заповнення пропусків даних) та операційного ризику страхових компаній.

2. На основі інтеграції розроблених модулів, а також систем *R* та *Eviews* побудовано архітектуру СППР для аналізу фінансових процесів у страхуванні.

3. Запропоновано процедуру адаптації виродженої статистичної вибірки з метою подальшої побудови УЛМ. Встановлено, що теорія екстремальних значень є ефективним способом обробки екстремальних значень у актуарних процесах. Експериментально доведено наближення використаних даних до класу узагальненого розподілу екстремальних значень із початковим порогом 6,65. Для оцінювання параметрів успішно застосовано байєсівський підхід з використанням апіорних та апостеріорних розподілів параметрів.

4. Побудовано нові узагальнені лінійні моделі для обраних актуарних процесів, які забезпечують оцінювання короткострокових прогнозів страхових виплат високої точності. Встановлено, що кращою є модель з гамма розподілом та логарифмічною функцією зв'язку. Для побудованих моделей відносні похибки оцінок прогнозів знаходяться в межах 5 – 12%.

5. Розроблено модель у формі мережі Байєса для ймовірнісного оцінювання операційних ризиків страхових компаній. Функціонування такої мережі апробовано на прикладах з використанням фактичних статистичних даних – рейтингів СК України за договорами страхування життя у період 2003-2016 рр. Ризик банкрутства СК у випадку 25% ймовірності настання страхового випадку та 100% виплати страхової премії складає 87% при тому, що 78,2% страхувальників вчасно здійснюють платежі за договорами про страхування.

Результати дослідження впроваджено у навчальний процес кафедри математичних методів системного аналізу ННК «Інститут прикладного системного аналізу» Національного технічного університету «Київський політехнічний інститут імені Ігоря Сікорського»; розроблена система використовується в умовах реального середовища існування страхової компанії Аркада-Гарант.

### **Особистий внесок здобувача**

Усі результати, що складають основний зміст роботи, отримані автором самостійно. У друкованих працях, написаних у співавторстві, здобувачеві належать: аналіз методів побудови математичних моделей для актуарних процесів, створення багатокрокової методології із використанням математичного апарату узагальнених лінійних моделей [1]; аналіз методів наближення експериментальних даних до класу узагальнених розподілів екстремальних значень, вибору оптимального порогу, дослідження методики аналізу екстремальних значень для оцінювання невідомих параметрів узагальнених лінійних моделей [3]; розробка структури мережі Байєса для оцінювання ризику актуарного процесу [2]; побудова узагальнених лінійних моделей для прогнозування актуарних процесів [6].

### **Апробація результатів роботи**

Основні наукові результати і положення дисертаційної роботи представлялись, доповідались і обговорювались на міжнародних конференціях, семінарах і публікувались у збірках наукових праць з метою апробації поточних результатів досліджень: «Міжнародна наукова студентська конференція МНСК-2013», 12 – 18 квітня 2013 р., м. Новосибірськ, Росія [1]; Інформаційні процеси та технології «Інформатика – 2013», 22 – 26 квітня 2013р., СевНТУ [2]; «Інтелектуальний аналіз інформації ІАІ-2013», 15-17 травня 2013 р., НТУУ «КПІ» [3]; «Міжнародна наукова студентська конференція МНСК-2014», 11 – 18 квітня 2014 р., м. Новосибірськ Росія [4]; Інформаційні процеси та технології «Інформатика – 2014», 22 – 26 квітня 2014р., СевНТУ [5]; "Информатика, управление и искусственный интеллект", 26 – 27 листопада 2014 р, м. Харків [6]; «Інформаційні технології, економіка та право: стан та перспективи розвитку», 24 – 25 вересня 2015 р., м. Чернівці [7]; «Математичне та програмне забезпечення інтелектуальних систем», 18 – 20 листопада 2015 р., м. Дніпропетровськ [8]; Congress of Information Technology, Computational and Experimental Physics (CITCEP 2015), 18 – 20 грудня 2016 р., Краків, Польща [9]; 20-й Міжнародний молодіжний форум «Радиоэлектроника и молодежь в XXI веке», 19 – 21 квітня 2016 р., м. Харків [10]; «Комп'ютерне моделювання та інформаційні технології в освіті» (КМІТО'2016), 19 – 20 травня 2016, м. Кривий Ріг [11]; Міжнародна науково-практична конференція «Проблеми і тенденції розвитку

сучасної економіки в умовах інтеграційних процесів», 11–12 листопада 2016 р., м. Херсон [12].

### **Публікації**

За матеріалами дисертаційного дослідження опубліковано 21 наукову працю, у тому числі 9 статей у провідних фахових виданнях (з них 2 у іноземних виданнях та 3 у виданнях України, що реферуються міжнародними наукометричними базами даних), 12 тез доповідей у матеріалах міжнародних і національних конференцій.

### **Структура дисертаційної роботи**

Дисертаційна робота складається із вступу, переліку умовних позначень, чотирьох основних розділів, висновків, списку використаних джерел (90 найменувань) та чотирьох додатків. Загальний обсяг роботи 182 сторінки. Основна частина дисертації містить 176 сторінки, в тому числі 42 рисунки, 12 таблиць.

## **ОСНОВНИЙ ЗМІСТ РОБОТИ**

У вступі обґрунтовано актуальність розробки нових математичних моделей та інформаційної системи підтримки прийняття рішень (СППР) для розв'язання задачі прогнозування величини страхових виплат у разі настання страхового випадку і оцінювання операційного ризику СК; визначені мета, об'єкт, предмет і методи дослідження, достовірність отриманих результатів, зв'язок з науковими програмами, планами, темами, наведено наукову новизну та практичне значення одержаних результатів, перелік публікацій за темою роботи, висвітлено особистий внесок здобувача.

У першому розділі виконано аналіз задач фінансового аналізу СК, кругообігу основних фінансових процесів у страхуванні. У цьому секторі економіки циркулюють значні грошові потоки, а страхові організації виступають активними учасниками перерозподілу валового внутрішнього продукту. Встановлено, що сьогодні у світі виконуються дослідження актуарних процесів у страхуванні, а саме, розв'язуються задачі оцінювання фінансових ризиків у страхуванні за допомогою економічних методів та методів експертного оцінювання. Однак, невідомі дослідження з використанням байєсівського підходу і побудови ймовірного висновку на основі неповних даних із врахуванням як безпосередніх залежностей рівня ризику від факторів ризику, так і залежності між факторами ризику. Виконано огляд моделей, які ґрунтуються на аналізі наслідків; моделей, які ґрунтуються на факторах ризику; регресійних та моделей байєсівського типу; виконано формалізований опис задачі оцінювання ризику СК. Встановлено, що на сьогодні недостатньо висвітлено питання побудови та оцінювання УЛМ і мереж Байєса (МБ). Існує необхідність побудови адаптивних алгоритмів, які можуть бути застосовані для обробки вироджених статистичних даних у пакетному режимі та у реальному часі.

Встановлено, що для розв'язання задачі прогнозування страхових виплат та оцінювання операційного ризику СК можна скористатись УЛМ та мережами Байєса. Моделі такого типу відрізняються простотою оцінювання їх структури і дають можливість враховувати залежності рівня ризику від факторів ризику і залежності між факторами ризику.



**Другий розділ** присвячено дослідженню двох класів моделей: УЛМ і байєсівських мереж. УЛМ є розширенням класу лінійних моделей. Досліджуваний вектор  $y$  має  $n$  компонент, який передбачає реалізацію випадкової незалежної величини  $Y$  з середнім  $\mu$ . Систематична частина моделі визначається вектором  $\mu$  з невідомими параметрами  $\beta_1, \beta_2, \dots, \beta_p$ . Узагальненою лінійною моделлю називають модель виду:

$$y = g^{-1} \left( \sum_{i=1}^n \beta_i g_i(x) \right),$$

де  $g(y)$  – функція зв'язку. Зазвичай припускається, що залежна змінна має експоненціальний розподіл. Характеристика УЛМ передбачає знання: (1) закону розподілу залежної змінної; (2) характеристик та параметрів функції зв'язку  $g(\cdot)$ ; (3) характеристику лінійного предиктора  $(X \cdot \beta)$ .

Функція зв'язку пов'язує лінійний предиктор  $\eta$  із значенням оцінки  $\mu$  величини  $Y$ . В класичній лінійній моделі середнє значення та лінійний предиктор є ідентичними;  $\eta$  і  $\mu$  вибираються довільним чином із множини дійсних чисел. Однак, при використанні розподілу Пуассона обов'язковою вимогою є  $\mu > 0$ . Оскільки моделі ґрунтуються на незалежних крос-класифікаційних даних із мультиплікативним ефектом, то функція зв'язку виражається через логарифмічну функцію:  $\eta = \log(\mu)$ ; тут  $\mu$  – обов'язково додатне. Таким чином, функція зв'язку є стохастичною, але з систематичними компонентами, які пов'язані функцією  $\theta$ . Функція зв'язку в загальному вигляді представляється так:

$$g(\mu) = \theta = X\beta,$$

$$g^{-1}(g(\mu)) = g^{-1}(\theta) = g^{-1}(X\beta) = \mu = E(y).$$

**Залишки в УЛМ.** Для нормально розподіленої моделі залежна змінна представляється у вигляді:  $y = \bar{\mu} + (y - \bar{\mu})$ , тобто ця величина виражається сумою прогнозного значення та залишків. Залишки використовуються для дослідження адекватності моделі, вибору функції зв'язку, аналізу дисперсії та елементів лінійного предиктора. УЛМ потребують розширеного означення залишків з метою їх застосування до розподілів різної природи. Залежно від форми розподілу визначено три види узагальнених залишків: Пірсона, Енскомба і залишки відхилення. Також досліджено типи розподілів у актуарному моделюванні, які включені до узагальненого класу розподілів: нормальний, розподіл Пуассона, біноміальний та гамма розподіли.

Розглянуто особливості побудови мереж Байєса та висновків на їх основі. Формально МБ – це трійка  $N = \langle V, G, J \rangle$ , де  $V$  – множина змінних;  $G$  – спрямований ациклічний граф, вузли якого випадкові змінні моделюючого процесу;  $J$  – розподіл ймовірностей змінних  $V = \{X_1, X_2, \dots, X_n\}$ . Для множини змінних повинна виконуватись марковська умова – кожна змінна не залежить від усіх інших змінних, за винятком батьківських попередників цієї змінної. Для опису МБ необхідно визначити структуру моделі у вигляді спрямованого ациклічного графа і параметри умовних розподілів ймовірностей (УРЙ).

**Третій розділ** присвячено розробці методики побудови математичних моделей для прогнозування актуарних процесів з використанням СППР. У роботі використовується структурний метод побудови математичних моделей з використанням системного підходу. Архітектура СППР подана на рис. 1.



Рис. 1. Архітектура СППР

Запропонована методика побудови математичної моделі актуарного процесу складається з етапів, поданих нижче.

**1 Етап:** *Системний аналіз досліджуваного процесу* на основі експертних оцінок, візуального дослідження вимірів вхідних і вихідних змінних, вивчення існуючих моделей та іншої доступної інформації. На першому етапі виконується статистичний аналіз початкової вибірки з метою виявлення факторів виродженості даних та подальшого оцінювання структури моделей. У межах дисертаційної роботи досліджуються два класи моделей: УЛМ та мережі Байеса.

**2 Етап:** *Попередня обробка експериментальних даних.* Мета попередньої обробки даних полягає у відсіюванні грубих похибок і оцінюванні достовірності експериментальних результатів, перевірка відповідності результатів вимірювання одному із законів множини експоненціальних розподілів.

Процес попередньої обробки експериментальних (статистичних) даних, включає такі операції: (1) нормування та візуальний аналіз даних і, за необхідності, їх корегування; тобто їх логарифмування або приведення до зручного діапазону відповідних змінних; (2) корегування даних полягає у заповненні пропусків та зменшенні викидів (екстремальних значень), що виходять за основний діапазон значень змінних; (3) формування перших або різниць вищих порядків, які необхідні для аналізу відповідних складових часового ряду.

**3 Етап:** *Аналіз часових рядів на стаціонарність та можливу наявність нелінійності за допомогою статистичних критеріїв якості.*

Для розв'язання задачі, пов'язаної з визначенням наявності нелінійності у досліджуваному процесі, їх характеру використовують візуальний аналіз даних та формальні тести. За допомогою візуального аналізу виявляють існування ділянок з

лінійним або нелінійним трендом, в якійсь мірі наявність гетероскедастичності та значних викидів, які можуть суттєво впливати на якість моделі [8]. Існує також ряд формальних тестів на наявність нелінійності. У роботі використано тест Фішера для визначення наявності нелінійності.

**4 Етап:** *Обробка екстремальних значень.* Для аналізу екстремальних значень використано сучасну теорію екстремальних значень (ТЕЗ). Математичну модель екстремальних значень (ЕЗ) представлено у вигляді:

$$M_n = \max \{X_1, \dots, X_n\},$$

де  $X_1, \dots, X_n$  – послідовність незалежних випадкових величин з функцією розподілу  $F$ ;  $M_n$  – максимум досліджуваного процесу на інтервалі часу  $n$  з розподілом:

$$\Pr\{M_n \leq z\} = \Pr\{X_1 \leq z, \dots, X_n \leq z\} = \Pr\{X_1 \leq z\} \times \dots \times \Pr\{X_n \leq z\} = \{F(z)\}^n.$$

Функція  $F$  розглядається як наближена оцінка для  $F^n$ . Якщо послідовність констант  $\{a_n > 0\}$  та  $\{b_n > 0\}$  таких, що

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \rightarrow F(a_n x + b_n x)^n \rightarrow G(z),$$

при  $n \rightarrow \infty$ , то  $G$  – невідроджена функція розподілу, яка належить до одного з розподілів ЕЗ, наприклад, до узагальненого розподілу екстремальних значень (УРЕЗ):

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$

де  $\mu$  – параметр розподілу;  $\sigma$  – параметр масштабу;  $\xi$  – параметр форми розподілу. У роботі використано такі типи таких розподілів ЕЗ:

1) розподіл Гумбела:

$$G(z) = \exp\left\{-\exp\left(-\left(\frac{z-b}{a}\right)\right)\right\}, \quad -\infty < z < \infty;$$

2) розподіл Фреше:

$$G(z) = \begin{cases} 0, & z \leq b; \\ \exp\left(-\left(\frac{z-b}{a}\right)^{-\alpha}\right), & z > b; \end{cases}$$

3) розподіл Вейбулла:

$$G(z) = \begin{cases} \exp\left(-\left(-\left(\frac{z-b}{a}\right)\right)^\alpha\right), & z < b. \\ 1, & z \geq b \end{cases}$$

4) узагальнений розподіл Парето (УРП):

$$G(y, \sigma, \xi) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)_+^{-1/\xi},$$

Для всіх трьох випадків  $a > 0$ ,  $b$  – дійсне число. Для другої та третьої функцій параметр  $\alpha > 0$ .

**5 Етап:** *Визначення структури моделей-кандидатів передбачає такі кроки:* (1) – ідентифікація стохастичної і систематичної складових та функції зв'язку; (2) – формулювання припущення стосовно випадковості, систематичності та вигляду функції зв'язку; (3) – обчислення описових статистик та визначення величини відповідності побудованої моделі даним; (4) – визначення значущості предикторів за допомогою статистики Вальда та статистики міток; (5) – оцінювання характеристик елементів структури математичної моделі (залишків).

**6 Етап:** *Вибір методу оцінювання невідомих параметрів математичних моделей вибраних структур.* У роботі використано зважений метод найменших квадратів (УМНК), метод максимальної правдоподібності (ММП), метод Монте-Карло, байєсівський підхід. Після оцінювання моделі оцінки випадкового процесу визначаються похибками:  $\hat{\varepsilon}(k) = e(k) = y(k) - \hat{y}(k)$ , що дає можливість виконати аналіз характеристик випадкового процесу  $\{\varepsilon(k)\}$ .

**7 Етап:** Обґрунтування та вибір кращої з оцінених моделей-кандидатів за допомогою множини статистичних критеріїв адекватності моделі. Для визначення адекватності моделі використано інформаційний критерій Акайке, тест Хосмера-Лемешова, статистика Дарбіна-Уотсона,  $F$ –статистика Фішера. При формуванні статистичного висновку використано фактор Байєса  $BF(i, j)$  – відношення апостеріорних ймовірностей до апіорних. Також, актуарію надається можливість прийняття рішення щодо доречності деталізованого фінансового аналізу з метою вчасного запобігання настанню банкрутства СК; аналізу ефективності виконання умов договорів та виплати страхових платежів і страхових виплат; прийняття управлінських рішень та аналізу наслідків їх реалізації; а також ймовірного прогнозування величини страхових виплат.

**Четвертий розділ** присвячено експериментальним дослідженням – адаптації статистичних даних до побудови УЛМ за допомогою запропонованого алгоритму обробки ЕЗ. Для обробки статистичного ряду з  $n$  незалежних, однаково розподілених змінних  $X_1, \dots, X_n$  застосовано таку послідовність дій.

1. Групування вибірок даних з  $n$  спостережень. Такі вибірки повинні містити від 50 до 100 значень.

2. Визначення максимуму  $Z_i$  для кожного блоку  $i$ .

3. Наближення кожного блоку максимумів до розподілу ЕЗ.

Після апроксимації узагальненим розподілом для кожного з річних максимумів розраховується функція квантилю [3, 4]:

$$z_p = \begin{cases} \mu - (\sigma / \xi) (1 - (-\log(1 - p))^{-\xi}), & \xi \neq 0; \\ \mu - \sigma \log(-\log(1 - p)), & \xi = 0; \end{cases}$$

Припускається, що  $y_p = -\log(1 - p)$  і квантиль-функція матиме вигляд:

$$z_p = \begin{cases} \mu - (\sigma / \xi) (1 - (y_p)^{-\xi}), & \xi \neq 0; \\ \mu - \sigma \log(y_p), & \xi = 0; \end{cases}$$

4. Виконується оцінювання параметрів моделі та розв'язується задача пошуку блоку оптимальної довжини. Остання зводиться до пошуку співвідношення між величинами відхилення та дисперсії.

5. Графічна перевірка наближення розподілу ЕЗ.

Для обґрунтування якості екстраполяції моделей використано графічний аналіз даних.

*Графік щільності розподілу.* В основі цього графіка лежить порівняння емпіричної та апроксимуючої функцій щільності розподілу. Абсциса точки на графіку щільності розподілів є емпіричною функцією розподілу, у яку замість аргументу підставляють дані з вибірки, а ордината – це теоретична функція розподілу, куди аналогічно замість аргументу підставляють дані із статистичної вибірки. За допомогою цього графіка на практиці часто вдається запобігти ефекту «виродженості».

*Графік квантилів (Q-Q plot).* Графік квантилів дає можливість оцінити ступінь довіри до параметричних моделей і визначається множиною значень:

$$\left\{ X_{k,n}, F^{-1}\left(\frac{n-k+1}{n}\right), k = 1, \dots, n \right\}.$$

Якщо модель надає прийнятне згладжування, то графік має лінійну форму.

6. Визначення порогу екстремального значення.

Для підвищення ефективності результату наближення екстремальних даних до одного з розподілів ЕЗ застосовують так звані порогові моделі. Задача вибору оптимального порога ідентична задачі визначення розміру блока. Обидві задачі спрямовані на визначення балансу між відхиленням та дисперсією. Низький рівень порогу призводить до порушень асимптотичної апроксимації, а високий рівень забезпечує велику дисперсію.

7. Оцінювання невідомих параметрів моделі.

Після визначення порогу виконується оцінка невідомих параметрів узагальненого розподілу Парето за методом максимальної правдоподібності.

**Опис та попередня обробка статистичних даних.** Експериментальне дослідження ефективності запропонованої методики виконано за допомогою фактичних даних. Об'єм вибірки склав 247 вимірів, які включають такі змінні: назва страхової компанії; грошовий еквівалент страхових виплат; статистичний рік; кількість договорів, які уклала страхова компанія; страхові платежі; кількість страхових випадків за рік. Основна залежна змінна – страхові виплати, яка відображає здійснення грошових переказів при настанні страхового випадку.

На рис. 2 відображено графік залежності страхових виплат від статистичного року. Різкі зміни величини «*Страхові виплати*» пояснюються коливаннями величини «*Кількість страхових випадків*» для відповідного періоду. У страховій діяльності часто трапляються не виплати страхових премій через відсутність коштів. Саме тому графік має різкі коливання з наближення до нуля, наприклад, на інтервалі часу 2007 та 2011 років.



Рис. 2. Динаміка страхових виплат страхових компаній за 2003-2016 рр.

На рис. 3 відображено значення описових статистик. Коефіцієнт асиметрії коливається в межах 2,839 до 8,664. Останнє свідчить про наявність «правого хвоста» в розподілі змінних. Оскільки параметр ексцесу має значення більше трьох, то розподіл є гостровершинним.

View	Proc	Object	Print	Name	Freeze	Sample	Sheet	Stats	Spec
				Q_CASES		Q_ARRANG		DAMAGES	CHARGES
Mean				248.9717		42819.83		2779.190	47154.60
Median				65.00000		2431.000		780.0000	13017.40
Maximum				3800.000		2241084.		49575.00	557884.0
Minimum				2.000000		6.000000		1.200000	469.8000
Std. Dev.				526.4061		181679.5		6121.650	81935.75
Skewness				4.241486		8.663552		4.644020	2.838897
Kurtosis				24.51794		94.69488		29.05691	12.32584
Jarque-Bera				5505.864		89621.68		7875.493	1226.855
Probability				0.000000		0.000000		0.000000	0.000000
Sum				61496.00		10576499		686459.9	11647185
Sum Sq. Dev.				68167427		8.12E+12		9.22E+09	1.65E+12
Observations				247		247		247	247

Рис. 3. Описові статистики початкових даних

Статистика Жак-Бера дозволяє перевірити гіпотезу про нормальний розподіл, шляхом порівняння *третього моменту* (коефіцієнта асиметрії –  $S$ ) та *четвертого моменту* (коефіцієнта ексцесу –  $K$ ) з моментами нормального розподілу, для якого  $S=0$ ,  $K=3$ . Так, статистика Жак-Бера для змінної «Страхові виплати» дорівнює 7875,493, а ймовірність має значення близьке до нуля. Тому, існує підстава відхилити нульову гіпотезу стосовно нормальності.

Попередній аналіз даних свідчить про сильну виродженість вибірки, яка проявляється у вигляді значних шумових складових при побудові моделі (рис. 4). Саме тому виникає потреба у обробці масивів екстремальних даних та прийняття рішення про доречність їх попереднього логарифмування.

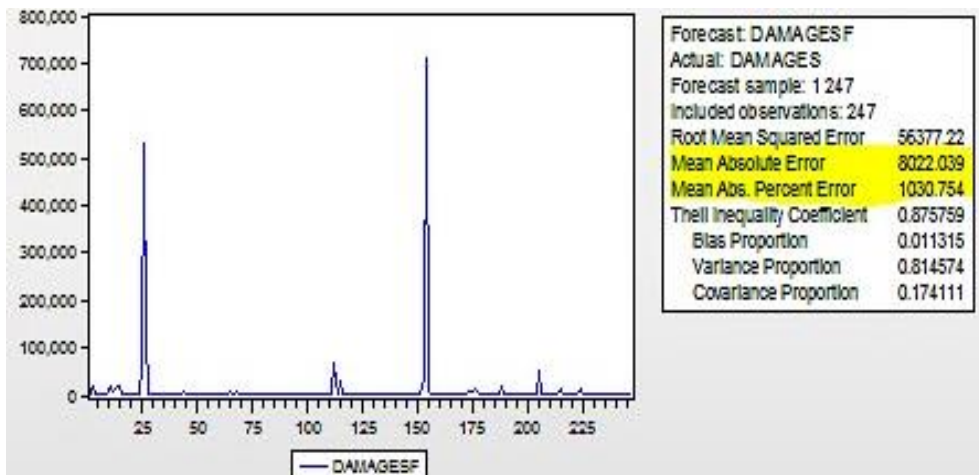


Рис. 4. Результати оцінювання моделі без попередньої обробки

Для вирішення задачі обробки вироджених статистичних даних запропоновано та теоретично обґрунтовано використання теорії ЕЗ. Аналіз описової статистики та візуальний аналіз логарифмованих даних дали можливість припустити про наближення даних до УРЕЗ або УРП. Далі розв'язується задачі оцінювання порогу для екстремальних значень.

Відносно високий поріг вибирається з метою зменшення зміщення, а з іншої сторони це означає, що лише декілька дослідів використовуються для оцінювання параметрів розподілу, тим самим гарантуючи збільшення оцінки дисперсії. Мета вибору величини порогу полягає в тому щоб уникнути зміщення моделі. Згідно розглянутого вище методу визначення величини порогу для експерименту прийнято значення 6,65. Графік Mean Residual Life Plot відображає залежність порогу від середнього залишку для оціненої моделі. Він слугує важелем перевірки вибраного порогу. Із рис. 5 видно, що після значення порогу 6 з'являються помітні відхилення від лінійності.

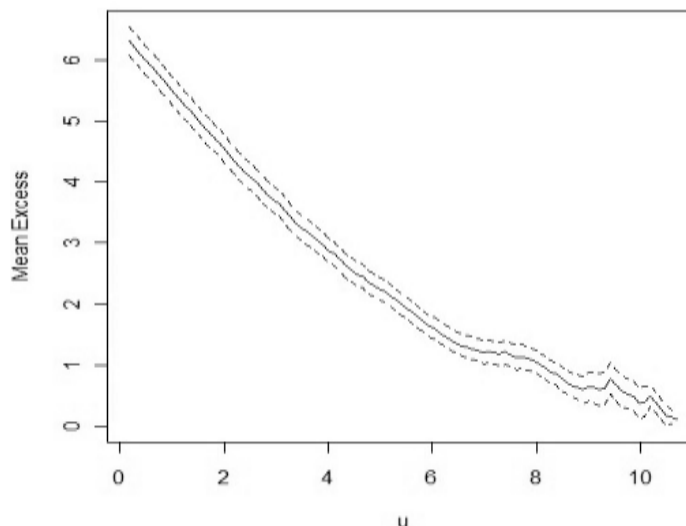


Рис. 5. Залежність значення порогу від середнього залишку УРП

Порівняльна характеристика параметрів розподілу представлена в таблиці 1. Вона показує, що кращим є наближення даних за допомогою УРП із незначною

похибкою та максимальним наближенням емпіричної кривої до теоретичної функції щільності розподілу.

Таблиця 1 – Порівняльна характеристика параметрів розподілів

Тип розподілу	Sigma		Xi		Логарифм правдоподібності	Перевищення порогу (за рік)	Кількість перевищень порогу
	ММП	СП	ММП	СП			
УРЕЗ	1,953	0,712	-0,650	0,095	487,812	—	—
УРП	0,777	0,346	-0,541	0,206	146,369	183,364	124

Оцінки параметрів побудованої моделі за допомогою байєсівського підходу зображено на рис. 6. Порівнюючи графіки щільності розподілу для побудованої моделі та наявної вибірки даних, можна помітити значні покращення моделі у термінах належності до одного з УРЕЗ.

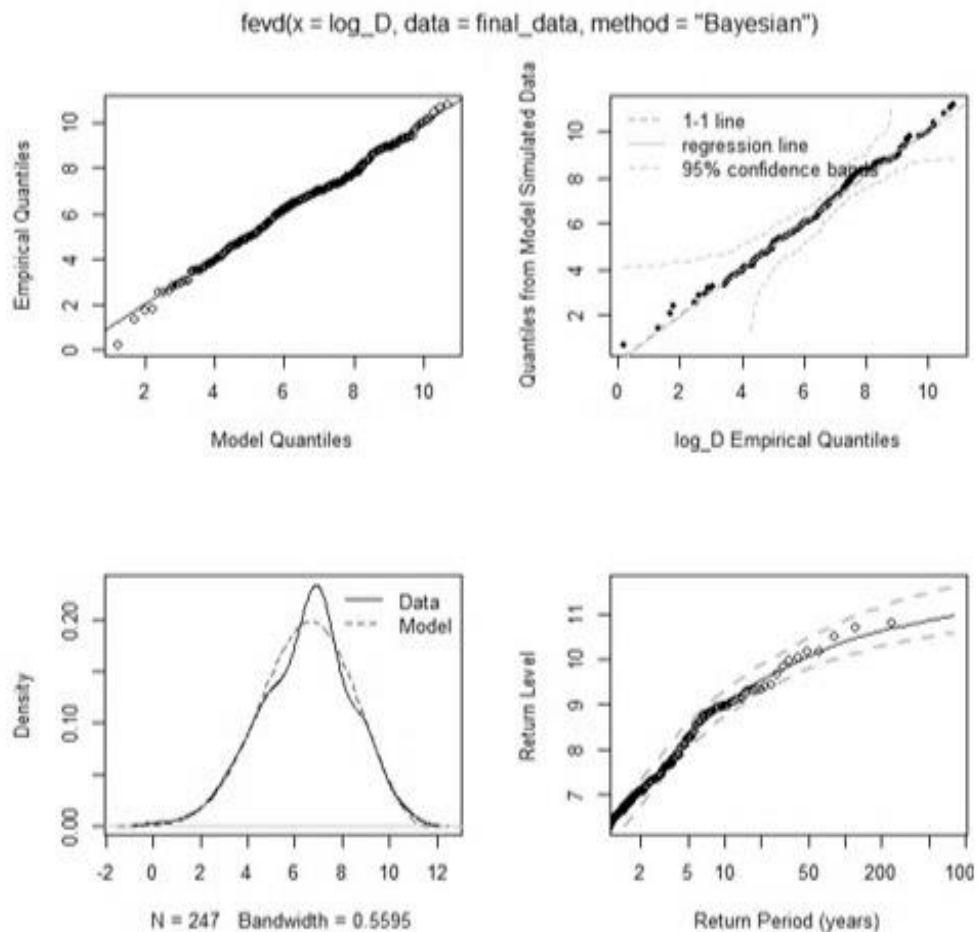


Рис. 6. Діагностика наближення моделі до одного з УРЕЗ

Порівнюючи отримані результати, слід зауважити, що байєсівський підхід демонструє кращі результати ніж метод максимальної правдоподібності та сприяє обґрунтованому вибору кращої моделі із запропонованих УРЕЗ. Таким чином, для оцінювання невідомих параметрів побудованих моделей, які належать до класу УРЕЗ, успішно використано байєсівський підхід з використанням апіорних та апостеріорних розподілів параметрів.



**Прогнозування актуарних процесів за допомогою УЛМ.** Для розв'язання задачі прогнозування величини страхових виплат СК використано річні дані за 2003-2015 рр. та трьох кварталів 2016 р.: страхові виплати, страхові платежі, кількість страхових договорів, частота появи страхових випадків, темпи зростання капіталу СК. В результаті застосування запропонованої методики побудовано 6 узагальнених лінійних моделей. Вони відрізняються розподілами залежної змінної (страхові виплати) та видом функції зв'язку. Встановлено, що мінімальні значення критерію Акаїке мають моделі з гамма або нормальним розподілами залежної змінної та логарифмічною функцією зв'язку – 23,84 і 25,30 відповідно. Для цих моделей показники відносної похибки співпадають і становлять 3,14%, що свідчить високу точність прогнозних значень страхових виплат. Критерій Байєса з шести моделей виділяє дві адекватні моделі – з гамма розподілом та логарифмічною функцією зв'язку і модель з нормальним розподілом та логарифмічною функцією зв'язку. Мінімальні значення критерію Байєса належать до гамма моделі з логарифмічною функцією зв'язку. Ця модель забезпечила найкраще наближення прогнозів до реальних даних з відносною похибкою 3,14%.

Після оцінювання втрат, пов'язаних із страховими виплатами за полісами страхування, наступною важливою задачею аналізу актуарних процесів є оцінювання операційного ризику. Для розв'язання цієї задачі побудовано модель у формі мережі Байєса. Для побудови МБ використано такі змінні: прихована змінна роки (Years), яка не є спостережуваною та доступною для навчання; кількість страхових договорів (Ins\_arrangements), кількість страхових випадків (Ins\_cases), страхові платежі (Ins\_charges), страхові виплати (Ins\_damages), ризик банкрутства (Ins\_loss).

Для оцінювання структури БМ у вигляді спрямованого ациклічного графа визначено такі причинно-наслідкові зв'язки між зазначеними вище змінними:

1)  $A \rightarrow \{C, Ch\}$ : укладення страхового договору (A) у статистичному році (Y) є причиною виникнення двох незалежних подій – внесення страхових платежів (Ch) клієнтом та настання страхового випадку (C), згідно відповідного договору страхування (A).

2)  $\{C, Ch\} \rightarrow D$ : СК, як гарант виконання умов договору (A) у разі настання страхового випадку (C) та при умові 100% внесення платежів (Ch) повинна здійснювати страхові виплати (D).

3)  $D \rightarrow L$ : відшкодування страхових виплат приносять збитки СК – виникає операційний ризик (L).

Тобто, змінна A – є причиною дії змінних C та Ch, наслідком дії яких є змінна D та L. Отже, множина змінних {A, Ch, C} є причинними, а {D, L} – наслідкові.

Структуру МБ довіри для оцінювання операційного ризику та ймовірності виплати грошової премії СК у разі виникнення страхового випадку відображено на рис. 7.

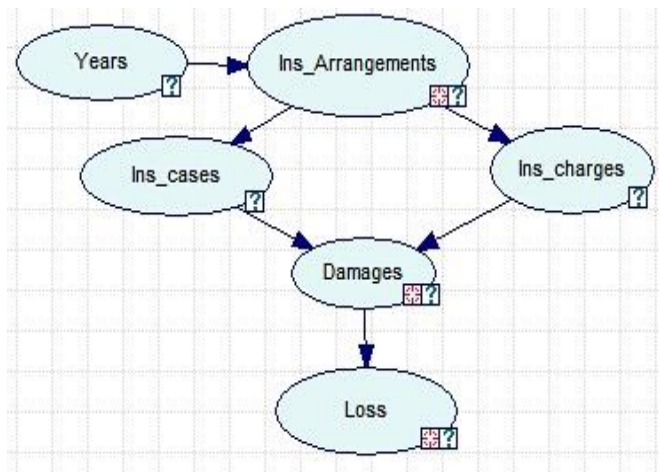


Рис. 7. Графічне представлення МБ довіри для ймовірнісного оцінювання актуарних процесів

Функціонування такої мережі апробовано на прикладах з використанням фактичних даних, а саме – рейтингів СК України за договорами страхування життя у період 2003-2016 рр. Результати оцінювання параметрів МБ відображено на рис. 8.

Ranked Targets	Probability	
Loss:PresentRisk	0.865	<div><div></div></div>
Ins_charges:PaidCharge	0.756	<div><div></div></div>
Ins_cases:AbsentCase	0.750	<div><div></div></div>
Damages.y2005	0.601	<div><div></div></div>
Ins_cases:PresentCase	0.250	<div><div></div></div>
Ins_charges:NotPaidCharge	0.244	<div><div></div></div>
Loss:AbsentRisk	0.135	<div><div></div></div>
Damages.y2006	0.058	<div><div></div></div>
Damages.y2004	0.056	<div><div></div></div>
Damages.y2009	0.047	<div><div></div></div>
Damages.y2003	0.045	<div><div></div></div>
Damages.y2011	0.039	<div><div></div></div>
Damages.y2010	0.034	<div><div></div></div>
Damages.y2007	0.029	<div><div></div></div>
Damages.y2008	0.029	<div><div></div></div>
Damages.y2014	0.027	<div><div></div></div>
Damages.y2012	0.023	<div><div></div></div>
Damages.y2013	0.013	<div><div></div></div>

Рис. 8. Результати оцінювання параметрів МБ

За допомогою МБ встановлено, що ризик банкрутства СК у випадку 25% ймовірності настання страхового випадку та 100% виплати страхової премії складає 87% при тому, що 78,2% страхувальників вчасно здійснюють платежі за договорами про страхування. Значна величина ризику банкрутства СК свідчить про відсутність ефективного механізму управління коштами, як власного капіталу, так і отриманих від страхових договорів. Динаміка оцінювання страхових виплат у випадку настання страхового випадку при 100% сплаті страхових платежів має вигляд, поданий на рис. 9.

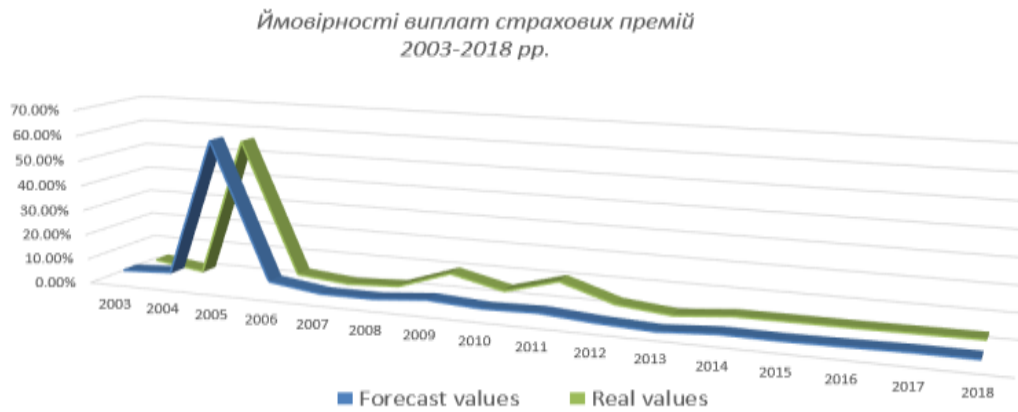


Рис. 9. Динаміка оцінювання страхових виплат.

З графічного представлення динаміки страхових виплат видно, що прогностні значення та апіорні ймовірності відрізняються не суттєво завдяки логічній структурі мережі Байєса та високій точності обчислень у мережі.

За допомогою спрощеної структури МБ обчислена величина операційного ризику СК при 33,9% або 44,2% виплаті страхової премії у разі настання (або відсутності) страхових випадків за договорами за якими поступали / не поступали регулярні платежі. Порівняльна таблиця результатів наведена в таблиці 2.

Таблиця 2 – Результати оцінювання спрощеної МБ

CASES	Present insurance case and Paid Charges	Absent insurance case and NotPaid Charges
Pr(Loss)	0,881	0,5
1-Pr(Loss)	0,119	0,5
Pr(Damages)	0,339	0,442

Із таблиці 2 видно, що ризик є високим навіть для випадку з 34% виплатою страхових премій СК та 100% вчасною сплатою платежів. Рівнозначні значення ризику отримано для протилежного випадку – не було страхового випадку і не сплачено платежі вчасно сплачені, тобто договір було розірвано. У такому випадку МБ відображає «відсутність ризику» на заданій множині станів, адже величина виплати премій досить низька. Таким чином, МБ – це потужний і ефективний математичний інструмент дослідження та відтворення реальної картини процесів у СК.

## ВИСНОВКИ

У ході виконання представленого дослідження запропоновано методику побудови узагальнених лінійних моделей для аналізу та прогнозування актуарних процесів. За результатами виконання роботи можна зробити такі висновки:

1. Виконано огляд моделей, які основані на аналізі наслідків; моделей, які основані на факторах ризику; моделі регресійного аналізу і функціональних кореляцій та моделі байєсівського типу. Встановлено, що існують методики побудови моделей опису та оцінювання актуарних процесів, але на сьогодні не достатньо висвітлено питання побудови та оцінювання УЛМ і мереж Байєса.

2. Запропоновано процедуру адаптації виродженої статистичної вибірки з метою подальшої побудови УЛМ. Встановлено, що теорія екстремальних значень є ефективним способом обробки екстремальних значень в фінансових процесах. Експериментально доведено наближення даних до класу УРЕЗ з початковим порогом 6,65. Для оцінювання параметрів успішно застосовано байєсівський підхід з використанням апіорних та апостеріорних розподілів параметрів, а також алгоритмів вибору кращої моделі.

3. Побудовано нові узагальнені лінійні моделі для обраних актуарних процесів, які забезпечують оцінювання короткострокових прогнозів стосовно страхових виплат високої точності. Обчислено оцінки прогнозів страхових виплат за допомогою побудованих у роботі узагальнених лінійних моделей. Встановлено, що кращою є модель з гамма розподілом та логарифмічною функцією зв'язку, результат котрої отримано за 4 ітерації алгоритму оцінювання параметрів моделі. При цьому БІК отримав значення 0,048.

4. Розроблено мережу Байєса для ймовірнісного оцінювання операційних ризиків страхових компаній. Функціонування мережі апробовано на фактичних статистичних даних, а саме – рейтингів СК України по договорах страхування життя у період 2003 – 2016 рр. Ризик банкрутства СК у випадку 25% ймовірності настання страхового випадку та 100% виплати страхової премії складає 87% при тому, що 78,2% страхувальників вчасно здійснюють платежі за договорами про страхування. Значна величина ризику банкрутства СК свідчить про відсутність ефективного механізму управління коштами, як власного капіталу, так і надходжень, отриманих від страхових договорів.

5. Проаналізовано якість побудованих моделей у формі УЛМ з використанням множини статистичних критеріїв. Встановлено, що такий підхід до моделювання є перспективним, оскільки модель демонструє високі показники якості. Відносні похибки оцінок прогнозів є незначними і коливаються в межах 5 – 12% для всіх моделей, а за допомогою інформаційного Критерію Акайке та Байєса обґрунтовано та обрано кращу модель. У результаті порівняння результатів побудованих УЛМ з методом групового урахування аргументів (МГУА) отримано значні розбіжності у прогнозних значеннях та величина відносної похибки становить 49,18%.

6. На основі інтеграції розроблених модулів, а також модулів систем R та Eviews створено СППР для аналізу фінансових процесів у страхуванні. Вона дає можливість використовувати її через різні інтерфейси, зокрема через модуль Rcmdr середовища програмування R. Ключовою перевагою розробленої СППР є забезпечення високої точності узагальнених лінійних моделей.

## СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Дубініна С. В., Бідюк П. І. Методика побудови математичних моделей актуарних процесів / С. В. Трухан, П. І. Бідюк // Східноєвропейський журнал передових технологій, №14 (73), 2015 р. *Міжнародні наукометричні бази: Index Copernicus, BASE, WorldCat, DOAJ, Citefactor, CrossRef, DRJI, PИНЦ. Автором розроблено та представлено багатокрокову методологію моделювання з використанням математичного апарату узагальнених лінійних моделей.*
2. Дубініна С. В., Бідюк П. І. Застосування мереж Байєса до побудови моделей оцінювання ризику актуарних процесів / С. В. Трухан, П. І. Бідюк // Scientific Journal ScienceRise, Т. 8, № 2 (25), 2016 р. *Міжнародні наукометричні бази: Index Copernicus, DOAJ, BASE, DRJI, Newton Library, Universia, SIS, GIF, OpenAire, PИНЦ. Автором розроблено та представлено структуру мережі Байєса для оцінювання ризику актуарного процесу.*
3. Дубініна С. В., Бідюк П. І. Методика аналізу екстремальних даних та її використання при оцінюванні параметрів узагальнених лінійних моделей / С. В. Трухан, П. І. Бідюк // Радіoeлектроніка, інформатика, управління, №1, 2016 р., м. Запоріжжя). *Міжнародні наукометричні бази: Thomson Reuters Web of Science (WoS), Index Copernicus, Academic Keys, Akademik Dizin, ARDI, COPAC, CrossRef, DOAJ, DOI, DRJI, EBSCO, Google Scholar, Jour Informatics, MIAR, OAJI, Open Science Directory, PUBGET, Sciary. Автором розроблено та представлено методику аналізу екстремальних значень на основі вибору оптимального порогу та оцінювання невідомих параметрів узагальнених лінійних моделей.*
4. Dubinina S., Bidyuk P. Application of Extreme Value Theory to Generation and Analysis of Pseudorandom Samples / Svitlana Trukhan, Petro Bidyuk // Journal of Mathematics and System Science, 6, 2016 p., 129-138 pp., USA. *(Іноземне видання.) Автором розроблено і представлено метод застосування теорії екстремальних значень до аналізу псевдовипадкових даних.*
5. Bidyuk P., Dubinina S., Systemic approach to estimation of financial risks // P. Bidyuk, S. Trukhan // ITHEA International Journal "Information Content and Processing (IJ ICP)" – V. 3, 2016 p., № 3, Bulgaria – www.ithea.org. *(Іноземне видання.) Автором представлено якісний аналіз невизначеностей у СППР, побудовано архітектуру СППР для аналізу та оцінювання операційних ризиків.*
6. Дубініна С. В., Бідюк П. І. Прогнозування актуарних процесів за допомогою узагальнених лінійних моделей / С. В. Трухан, П. І. Бідюк // Наукові вісті НТУУ «КПІ», №2, 2014 р. *Міжнародні наукометричні бази: WorldCat, Google Scholar, OpenAIRE, PИНЦ та EBSCO. Автором побудовано та представлено нові узагальнені лінійні моделі для прогнозування актуарних процесів.*
7. Дубініна С. В., Бідюк П. І. Оцінювання узагальнених лінійних моделей за байєсівським підходом в актуарному моделюванні / Бідюк П. І., Трухан С. В. // Наукові вісті НТУУ «КПІ», № 6, 2014 р. *Міжнародні наукометричні бази: WorldCat, Google Scholar, OpenAIRE, PИНЦ та EBSCO. Автором виконано та представлено порівняльний аналіз методів оцінювання узагальнених лінійних моделей.*
8. Згуровський М. З., Дубініна С. В., Бідюк П. І. Методика застосування теорії екстремальних значень для аналізу даних / М. З. Згуровський, С. В. Трухан, П. І. Бідюк // Наукові вісті НТУУ «КПІ», № 1, 2016 р. *Міжнародні наукометричні*

бази: WorldCat, Google Scholar, OpenAIRE, PИЦ та EBSCO. Автором розроблено та представлено методи обробки екстремальних значень, побудовано архітектуру СППР для аналізу та генерування псевдовипадкових величин.

9. Бідюк П. І., Дубініна С. В., Кожухівська О. А. Проектування і застосування інформаційно-аналітичної системи для аналізу і прогнозування фінансово-економічних процесів / П. І. Бідюк, С. В. Трухан, О. А. Кожухівська // Електронний науковий журнал “Системні науки і кібернетика” ННК “ІПСА” НТУУ «КПІ», № 1, 2013 р., <http://mmsa.kpi.ua/ssc>. Автором розроблено і представлено нові моделі обраних фінансово-економічних процесів.

10. Дубініна С. В. «Міжнародна наукова студентська конференція МНСК-2013», 12-18 квітня 2013 р., м. Новосибірськ. – “Обобщенные линейные модели как метод прогнозирования актуарных процессов”

11. Дубініна С. В. Інформаційні процеси та технології. «Інформатика – 2013», 22-26 квітня 2013р., м. Севастополь. – “Система підтримки прийняття рішень для моделювання та менеджменту актуарних процесів”.

12. Дубініна С. В. «Інтелектуальний аналіз інформації ІАІ-2013» XIII Міжнародна наукова конференція ім. Таран Т. А. Интеллектуальный анализ информации (ІАІ – 2013), 15-17 травня 2013 р., м. Київ. – “Прогнозування актуарних процесів за допомогою узагальнених лінійних моделей”.

13. Дубініна С. В. «Міжнародна наукова студентська конференція МНСК-2014», 11-18 квітня 2014 р., м. Новосибірськ. – “Методы генерации эталонных данных для СППР”.

14. Дубініна С. В. Інформаційні процеси і технології «Інформатика — 2014», VII Міжнародна науково-практична конференція, 22—26 квітня 2014 р., м. Севастополь – “Імітаційне моделювання як інструмент генерування модельних даних для СППР”.

15. Дубініна С. В. Науково-технічна конференція студентів, магістрантів та аспірантів, "Информатика, управление и искусственный интеллект", 26 – 27 листопада 2014 р, м. Харків. – “Байєсівська методологія як інструмент визначення значущості методів оцінювання якості моделей”.

16. Дубініна С. В. Міжнародній науково-практичній конференції «Інформаційні технології, економіка та право: стан та перспективи розвитку» (ІТЕП-2015), 24 – 25 вересня 2015 р., м. Чернівці – “Модель обробки екстремальних даних для побудови УЛМ”.

17. Дубініна С. В. XIII Міжнародній науково-практичній конференції “Математичне та програмне забезпечення інтелектуальних систем”, 18 – 20 листопада 2015 року, м. Дніпропетровськ – “Теорія екстремальних значень як інструмент моделювання випадкових процесів”.

18. Dubinina S., Bidyuk P. Congress of Information Technology, Computational and Experimental Physics (CITCEP 2015), December 18 – 20, 2015, Krakow (Poland). – “Systemic approach to estimation of financial risks”. Автором запропоновано етапи розробки та побудови методології системного аналізу для оцінювання фінансових ризиків.

19. Дубініна С. В. 20-й Юбилейный международный молодежный форум «Радиоэлектроника и молодежь в XXI веке», 19 – 21 квітня 2016 р., м. Харків –

“Застосування байєсівських мереж для оцінювання та прогнозування актуарних процесів”.

20. Дубініна С. В. VII Всеукраїнської науково-методичної конференції «Комп’ютерне моделювання та інформаційні технології в освіті» (КМІТО’2016), 19 – 20 травня 2016, м. Кривий Ріг – “Байєсівські мережі як інструмент оцінювання та прогнозування актуарних процесів”.

21. Дубініна С. В. Міжнародна науково-практична конференція «Проблеми і тенденції розвитку сучасної економіки в умовах інтеграційних процесів: теоретичні та практичні аспекти», 11– 12 листопада 2016 р., м. Херсон – “Системний підхід у задачах актуарного прогнозування та оцінювання операційних ризиків”.

## АНОТАЦІЇ

**Дубініна С. В. Байєсівські методи моделювання актуарних процесів та оцінювання ризиків страхових компаній. – Рукопис.**

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту. – Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, Київ, 2017 р.

Дослідження спрямоване на прогнозування величини страхових виплат у разі настання страхового випадку та оцінювання операційного ризику страхових компаній. Моделі розроблено у формі узагальнених лінійних моделей (УЛМ) та мережі Байєса, що є елементом новизни дослідження. Проаналізовано фактори, що впливають на страхові виплати, і виявлено існування закономірності зростання динаміки страхових виплат у залежності від кількості страхових випадків та страхових платежів. Запропоновано процедуру адаптації виродженої статистичної вибірки з метою подальшої побудови УЛМ. Експериментально доведено наближення даних до класу узагальнених розподілів екстремальних значень з початковим порогом 6,65. Для оцінювання параметрів успішно застосовано байєсівський підхід з використанням апіорних та апостеріорних розподілів параметрів, а також алгоритмів вибору кращої моделі. Побудовано нові узагальнені лінійні моделі для обраних актуарних процесів, які забезпечують оцінювання короткострокових прогнозів стосовно страхових виплат прийнятною точністю. Обчислено оцінки прогнозів страхових виплат за допомогою побудованих у роботі узагальнених лінійних моделей. Встановлено, що кращою є модель з гамма розподілом та логарифмічною функцією зв’язку. Також, для ймовірнісного оцінювання операційних ризиків страхових компаній розроблено мережу Байєса. Функціонування такої мережі апробовано на прикладах з використанням фактичних статистичних даних, а саме – рейтингів СК України за договорами страхування життя у період 2003-2016 рр. Ризик банкрутства СК у випадку 25% ймовірності настання страхового випадку та 100% виплати страхової премії складає 87% при тому, що 78,2% страхувальників вчасно здійснюють платежі за договорами про страхування. Отримані результати демонструють високу точність відповідно до загальноприйнятих статистичних критеріїв якості. Комплексна модель дає можливість вчасно запобігти банкрутства СК і якісно проаналізувати необхідну величину страхових виплат при настанні страхового випадку за договорами



страхування.

*Ключові слова:* актуарні процеси, математичне моделювання, узагальнені лінійні моделі, мережі Байеса, екстремальні значення та теорія екстремальних значень, розподіли екстремальних значень.

**Дубинина С. В. Байесовские методы моделирования актуарных процессов и оценивание рисков страховых компаний. – Рукопись.**

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.23 – системы и методы искусственного интеллекта. – Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», Киев, 2017 г.

Исследование направлено на прогнозирование величины страховых выплат в случае наступления страхового случая и оценивания операционного риска страховых компаний (СК). Модели разработаны в форме обобщенных линейных моделей (ОЛМ) и сети Байеса, что является элементом научной новизны. Решаемая задача имеет высокий приоритет, так как наиболее распространенный фактор банкротства СК наступает через нерегулируемый либо отсутствующий финансовый анализ деятельности компаний, который сопровождается невыплатой страховых премий. Проанализировано факторы, влияющие на страховые выплаты и выявлено закономерности изменения их динамики в зависимости от количества страховых случаев и платежей, поступающих за договорами страхования. Показано, что теория экстремальных значений – эффективный способ обработки экстремальных значений в финансовых процессах. Экспериментально доказано приближение данных к классу обобщенных распределений экстремальных значений с начальным порогом 6,65. Построенные новые ОЛМ для описания актуарных процессов обеспечивают оценки краткосрочных прогнозов высокой точности, касающихся страховых выплат. Оценки прогнозов страховых выплат вычислены с помощью построенных ОЛМ. Наилучшей оказалась модель с гамма распределением и логарифмической функцией связи. Для вероятностного оценивания операционных рисков страховых компаний построена сеть Байеса. Функционирование сети апробировано на примерах фактических данных, а именно – рейтингов СК Украины за договорами страхования жизни в период 2003 – 2016 гг. Риск банкротства СК в случае 25% вероятности наступления страхового случая и 100% выплаты страховых премий составляет 87% притом, что 78,2% страховщиков вовремя осуществляют платеж за договорами страхования. Существенный риск банкротства СК свидетельствует об отсутствии эффективного механизма управления денежными средствами. Построенная комплексная модель предоставляет возможность вовремя предотвратить банкротство СК и качественно проанализировать необходимую сумму страховых выплат при наступлении страхового случая за договорами страхования.

*Ключевые слова:* актуарные процессы, математическое моделирование, обобщенные линейные модели, сети Байеса, экстремальные значения и теория экстремальных значений.



Dubinina S. **Bayesian methods of simulating actuarial processes and risk estimation of insurance company.** – Manuscript.

Thesis in fulfillment of the requirements for the degree of candidate of engineering sciences on the specialty of 05.13.23 – Systems and methods of artificial intelligence. – National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, 2017.

The research is directed towards forecasting of insurance payments in respective cases and estimation of operational risk for insurance companies. The necessary mathematical models were developed in the form of generalized linear models (GLM) and Bayesian networks what is a new scientific element of the research. The problem being solved has a high priority due to the fact that the most known factor of companies' bankruptcy is induced by the non-settled or not performed at all financial analysis of insurance company activities that is accompanied by non-payments of insurance premiums and mathematical modeling is not actively used. The most important factors of influence on insurance payments were analyzed and the payment dynamics was revealed depending on the number of insurance cases and payments according to the insurance policies. An adaptation procedure based on extreme value theory was proposed for processing degenerated statistical samples for constructing GLM. It was shown that the extreme value theory is very effective basis for processing available financial processes outliers. It was also shown experimentally that the statistical data used were approaching to the class of generalized extreme values distribution with initial threshold of about 6.65. The threshold was estimated in this specific case by making use of visual data analysis and special statistical methodology. The model parameters were successfully estimated with application of Bayesian approach using prior and posterior parameters distributions as well as algorithms for selecting the best models. Each step of computations was accompanied by application of appropriate set of statistical quality criteria what led to high quality final results.

New generalized linear models were constructed for the selected actuarial processes that provide for a possibility of estimating high quality short term forecasts regarding insurance payments. It was established that the best model was the one with gamma distribution and logarithmic link function. To get the possibility for probabilistic estimation of operational risks the model in the form of Bayesian network was developed using a set of selected variables. The model in the form of directed acyclic graph with the variables as nodes was developed and tested with actual statistical data reflecting activities of Ukrainian insurance companies in the area of life insurance within the period of 2003 – 2016. Thus, the bankruptcy risk, in the case of 25% probability of insurance situation and payment of 100% of insurance premiums, exhibited 87% in conditions that 78.2% of clients timely perform their payments according to their policies. A noticeable value of insurance company bankruptcy reflects an absence of effective mechanism for companies' funds management: both companies' own means and regular premium payments. Thus, models in the form of Bayesian networks create a powerful instrument for studying actuarial processes and estimating possible financial losses. The experimental results generated demonstrate high accuracy with respect to the widely accepted statistical quality criteria. The complex model developed provides a possibility for timely prevention of insurance company bankruptcy and perform an analysis for necessary volume of insurance

payments according to the specific insurance policies. All computational experiments were performed with the decision support system designed and implemented by the author. The system constitutes a handy data analysis instrument for insurance company managers and can be easily extended with new functions.

*Keywords:* actuarial processes, mathematical modeling, generalized linear models, Bayesian networks, extreme values and extreme value theory, distributions of extreme values.